D-JEPA: DESIGN-RECOVERABLE JEPA REPRESENTATION WITH SWAPPABLE PHYSICS DECODERS

Anonymous authors
Paper under double-blind review

ABSTRACT

Joint-Embedding Predictive Architectures (JEPAs) provide a framework for learning compact representations without directly reconstructing high-dimensional observations. However, in parameterized physical systems, learned representations can entangle geometry with operating conditions and task-specific physical responses, limiting their reuse across prediction tasks. We introduce D-JEPA (Design-recoverable JEPA), a geometry-centric JEPA that computes a compact representation from geometry alone and reuses it across operating conditions and physical response spaces through lightweight physics-specific decoders. An explicit design-recoverability objective encourages the geometry latent to preserve information about the underlying design variables, enabling the representation to support design analysis and optimization. We further identify a case-level collapse failure mode in which target representations become nearly invariant across distinct geometries despite low reconstruction error, and mitigate it using case-level variation constraints and auxiliary target reconstruction. Across four 3D aerodynamic, hydrodynamic, and structural benchmarks, D-JEPA maintains or improves full-field prediction accuracy while achieving near-perfect linear recoverability of design parameters. The frozen geometry representation can be reused at held-out operating conditions and transferred to a structural response task with fewer trainable parameters. Finally, the representation supports differentiable design optimization, with designs validated using high-fidelity CFD, preserving the predicted ranking of candidate designs. These results demonstrate that separating a reusable geometry representation from physics-specific prediction provides a practical representation for scientific surrogate modeling and design.

1 INTRODUCTION

Many scientific and engineering problems require repeatedly solving parameterized partial differential equations (PDEs) as geometry, operating conditions, or boundary conditions vary. In aerodynamic design, Aerodynamic Databases (AeroDBs) map flight conditions and vehicle geometries to quantities of interest (QoIs) such as lift, drag, and structural loads, supporting applications from preliminary aircraft design and aeroelasticity to Certification by Analysis (CbA). However, populating an AeroDB requires evaluating a combinatorial space of geometries and operating conditions, making high-fidelity computational fluid dynamics (CFD) prohibitively expensive. The scale of this cost is evident in NASA's Space Launch System databases, which required thousands of CFD simulations and tens of millions of CPU-core hours (NASA Advanced Supercomputing Division, 2017; 2018), and the NASA X-57 aerodynamic database, which involved thousands of simulations with meshes containing hundreds of millions of cells (NASA Advanced Supercomputing Division, 2021; Frederick et al., 2025). These computational demands have motivated neural surrogates that amortize simulation cost by learning the underlying input–output mapping (Brunton et al., 2020). While recent surrogates can substantially accelerate CFD prediction, many downstream aerodynamic applications require only a small set of QoIs rather than complete high-dimensional flow fields, motivating reusable geometry representations that directly support physical prediction and downstream design tasks.

A large body of recent work has developed neural surrogates for predicting high-dimensional physical fields on complex engineering geometries. Transformer-, mesh- and geometry-aware architectures such as Transolver, DoMINO, GeoTransolver and AB-UPT have demonstrated accurate

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

prediction of physical responses on general engineering domains (Wu et al., 2024a; Ranade et al., 2025; Adams et al., 2025; Alkin et al., 2025). These approaches are particularly valuable when the complete physical field is required. However, aerodynamic design workflows also require repeated evaluation of integrated quantities such as lift, drag and moments across large design spaces. More fundamentally, a surrogate optimized for a particular physical response does not necessarily provide a representation that can be reused when operating conditions or the physical quantity of interest changes. This motivates learning a reusable geometry representation that can support multiple downstream physical prediction and design tasks beyond a single geometry-to-field mapping.

Joint-Embedding Predictive Architectures (JEPAs) provide a natural framework for learning such representations. Rather than reconstructing high-dimensional observations directly, JEPA learns predictive relationships between latent representations (Assran et al., 2023; Bardes et al., 2024). Related predictive representation methods have likewise demonstrated that useful information can be retained in compact latent spaces without explicit reconstruction (Grill et al., 2020; Baevski et al., 2022). For engineering surrogate modeling, however, the representation must remain useful as geometry, operating conditions and physical response spaces change. A representation that entangles these factors may be effective for a particular prediction task but difficult to reuse when the operating condition or physical quantity of interest changes. More broadly, self-supervised representation learning has identified representation collapse as a potential failure mode, motivating mechanisms such as redundancy reduction and variance preservation (Zbontar et al., 2021; Bardes et al., 2022; aron et al., 2021; Balestriero & LeCun, 2025). Recent work has begun to explore predictive latent representations for aerodynamic surrogate modeling. AeroJEPA, for example, predicts aerodynamic latent representations from geometry and operating conditions and demonstrates that such latent spaces can support linear probing and design-oriented operations (Giral et al., 2026). However, its representation remains tied to the aerodynamic prediction setting and operating conditions, leaving open a more general question: can geometry itself be represented independently of operating conditions and physical response, while retaining information that is explicitly recoverable for downstream design?

We propose D-JEPA (Design-recoverable JEPA), a joint-embedding architecture that separates geometry representation from physics prediction by encoding each geometry into a compact latent space and conditioning lightweight physics decoders on operating conditions. The geometry encoder receives neither operating conditions nor target physics, while an explicit design-recoverability objective encourages the latent to preserve geometric design information. This factorization enables the learned representation to be evaluated for predictive accuracy, design-variable recoverability and reuse across operating conditions and physical response spaces. We demonstrate these capabilities through full-field prediction, condition and cross-physics transfer, as well as differentiable design optimization validated with high-fidelity CFD. Our contributions are as follows:

1. A geometry-centered JEPA formulation for scientific surrogate modeling that explicitly factorizes a reusable geometry representation from operating-condition- and physics-specific prediction.
2. An explicit design-recoverability objective that encourages the shared geometry latent to preserve information about the underlying geometric design variables, making this information directly recoverable through a shallow probe.
3. A reusable geometry–physics factorization in which the geometry encoder is shared across operating conditions and physical response spaces, while lightweight physics-specific decoders can be swapped without retraining the geometry encoder.
4. Empirical evaluation of the resulting representation across parameterized aerodynamic benchmarks, operating-condition reuse, cross-physics transfer and differentiable design optimization with high-fidelity CFD validation.

2 METHOD

In this section, we present the mathematical formulation of D-JEPA, including the geometry representation, latent prediction, physics-specific decoding and design-recoverability objective.

2.1 PROBLEM FORMULATION

We consider the problem of predicting steady physical fields on a set of parameterized geometries under varying operating conditions and multiple sets of governing equations. Let a design be defined by its boundary surface $\partial\Omega$ embedded in $\Omega \subset \mathbb{R}^3$, represented in a canonical reference coordinate system. The surface is discretized as a point cloud $P = \{x_i\}_{i=1}^N$, with corresponding surface normals $N = \{n_i\}_{i=1}^N$. Let $c \in \mathbb{R}^{n_{\text{cond}}}$ denote the operating environment, such as Reynolds number, angle of attack, Mach number or structural load cases. We index different physical regimes by $s \in \mathcal{S}$, where each regime corresponds to a particular set of governing equations, such as fluid or structural physics. The corresponding ground-truth field is denoted by $F_s : \Omega \to \mathbb{R}^{d_s}$.

Each geometry is associated with a design vector $\theta \in \mathbb{R}^D$. The design parameters are never provided as inputs to the physics decoder; instead, they are used during training only through an auxiliary design-recoverability objective. The primary objective is to learn a geometry encoder E_ϑ that maps the surface representation to a compact latent representation

$$g = E_\vartheta(P, N), \qquad g \in \mathbb{R}^{d_g}, \tag{1}$$

where $d_g = 128$, together with lightweight physics-specific decoders $\{D_{\phi_s}\}$. Field prediction is given by

$$\hat{F}_s(q) = D_{\phi_s}(g, q, c), \qquad g = E_\vartheta(P, N), \tag{2}$$

where q denotes a query location. By structural construction, g does not receive c or s as inputs. Thus, operating-condition and physics-specific information can enter the prediction only through the downstream decoder, while the geometry encoder is shared across conditions and physics regimes. This provides the factorization underlying D-JEPA: geometry is represented once in g, while operating conditions and physical response spaces are handled by the corresponding decoder.

2.2 D-JEPA FRAMEWORK

The D-JEPA framework separates geometry g, operating conditions c and governing physics through the decoder ϕ_s, enabling design recoverability, reuse across operating conditions and transfer across physical response spaces (Figure 1). During training, the geometry encoder, target encoder, latent predictor, physics decoder and design probe are jointly optimized. The target encoder E_{target} receives an independently sampled ground-truth field point cloud and produces target tokens Z_t, while the latent predictor $f_{\text{pred},\theta}(g, c)$ produces predicted tokens $\hat{Z}_t$. Gradients from $\mathcal{L}_{\text{lat}}$ propagate through both the predictor and target encoder; no stop-gradient or EMA teacher is used. The target encoder is used only during training and is discarded at inference. The JEPA predictor provides an auxiliary latent prediction signal that shapes the geometry representation, whereas the physics decoder $D_{\phi_s}(g, q, c)$ is responsible for direct physical-field prediction. The two pathways therefore serve distinct roles. The JEPA pathway provides representation-level supervision by requiring the geometry-dependent prediction $\hat{Z}_t$ to match target tokens derived from the physical field, while the physics decoder maps the shared geometry representation directly to the physical field at arbitrary query locations. Thus, $\mathcal{L}_{\text{lat}}$ shapes the representation, whereas $\mathcal{L}_{\text{rec}}$ evaluates its utility for physical-field prediction.

2.3 COORDINATE-AWARE GEOMETRY ENCODING

A local point-attention mechanism can capture geometric features such as local curvature but does not by itself provide an explicit representation of global proportions and orientation. We augment the learned point-transformer representation (Wu et al., 2024b) with two parameter-free geometric descriptor blocks computed in the canonical reference frame. The first descriptor is a global descriptor s_{glob} with 15 dimensions. It contains the per-axis mean and standard deviation, bounding-box extent and the eigenvalues and off-diagonal entries of the 3×3 position covariance matrix. These features provide compact information about the global scale, extent and spatial distribution of the geometry. The second descriptor is a regional descriptor s_{reg} obtained from a $5 \times 5 \times 5$ voxel grid. Each voxel contains 13 statistics summarizing the mean and standard deviation of positions, normal statistics and occupancy, resulting in a 1625-dimensional descriptor. In parallel, a learned point-transformer branch tokenizes the point cloud into $M = 512$ tokens with width $d = 128$, following the hierarchical point-set construction of Qi et al. (2017). Mean and max pooling produce

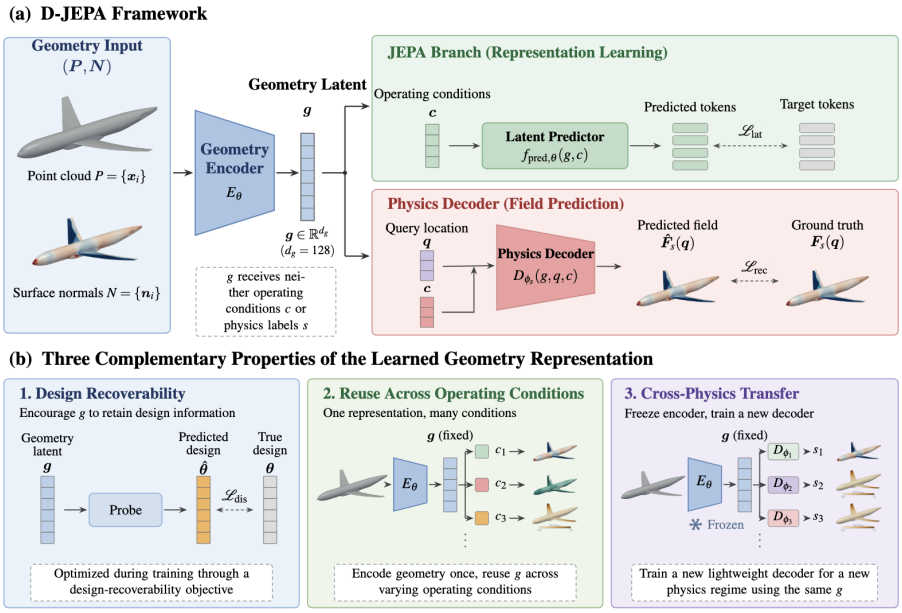


Figure 1: Overview of D-JEPA: (a) D-JEPA architecture and training pathways. (b) Design recoverability, condition reuse and cross-physics transfer.

$s_{\text{loc}} \in \mathbb{R}^{256}$. Layer sizes and neighbourhood parameters are given in Appendix B. The local, global and regional descriptors are concatenated, $[s_{\text{loc}}; s_{\text{glob}}; s_{\text{reg}}] \in \mathbb{R}^{1896}$, and passed through an MLP to obtain the geometry latent $g \in \mathbb{R}^{128}$. The resulting latent receives neither operating conditions c nor physics labels s, and the same geometry representation is used by every downstream physics decoder and operating point.

2.4 FiLM-Modulated Implicit Neural Field Decoder

For each physics regime s, the decoder maps a query location q to the corresponding local physical field and is modulated by the geometry latent g through Feature-wise Linear Modulation (FiLM) (Perez et al., 2018). This follows the implicit neural representation formulation used for continuous fields (Sitzmann et al., 2020; Mildenhall et al., 2021). The query location is encoded using a Fourier feature embedding with 16 octaves (Tancik et al., 2020), producing a 99-dimensional representation. The decoder additionally receives the local surface normal $n(q)$ and operating condition c. The decoder consists of four residual FiLM blocks. For each block l, the latent representation generates the modulation parameters

$$(\gamma_l, \beta_l) = \text{split}(A_l g), \tag{3}$$

and the hidden representation is updated as

$$h_{l+1} = h_l + \text{GELU}\left(\gamma_l \odot W_l \text{LayerNorm}(h_l) + \beta_l\right). \tag{4}$$

The predicted field is obtained from the final hidden representation as

$$\hat{F}_s(q) = W_{\text{out}} h_L. \tag{5}$$

This formulation allows the same geometry representation to condition different physics-specific decoders while operating conditions enter explicitly through the decoder. Automatic differentiation provides surrogate gradients with respect to the latent and design variables, without requiring a separate PDE-adjoint computation.

2.5 Training Objectives and Design Recoverability

During Stage 1, the geometry encoder, latent predictor, primary physics decoder and design probe are jointly optimized. A target encoder E_{target} receives an independently sampled ground-truth field

point cloud and produces target tokens

$$Z_t \in \mathbb{R}^{M \times d}. \tag{6}$$

The predictor maps (g, c) to predicted target tokens $\hat{Z}_t$. An auxiliary target decoder D_t reconstructs the physical field from Z_t, while the primary physics decoder D_{ϕ_s} predicts the field from the geometry representation g. The composite training objective is

$$\mathcal{L} = \mathcal{L}_{\text{lat}} + \lambda_r \mathcal{L}_{\text{rec}} + \lambda_d \mathcal{L}_{\text{dis}} + \lambda_s \mathcal{L}_{\text{sig}} + \lambda_c \mathcal{L}_{\text{sig-case}} + \lambda_v \mathcal{L}_{\text{var}} + \lambda_t \mathcal{L}_{\text{rec-t}}. \tag{7}$$

The latent prediction loss is $\mathcal{L}_{\text{lat}} = \left\| \hat{Z}_t - Z_t \right\|_2^2$ and the primary field reconstruction loss is $\mathcal{L}_{\text{rec}} = \mathbb{E}_q \left[\| D_{\phi_s}(g, q, c) - F_s(q) \|_2^2 \right]$. To explicitly encourage the geometry representation to preserve information about the underlying design, we introduce a design-recoverability loss

$$\mathcal{L}_{\text{dis}} = \| \text{probe}(g) - \theta \|_2^2 . \tag{8}$$

The design parameters are never supplied to the physics decoder. Instead, they supervise the representation through $\mathcal{L}_{\text{dis}}$, explicitly encouraging g to retain design-relevant information. Unlike a purely post-hoc probing analysis (Alain & Bengio, 2017), design recoverability is therefore directly optimized during representation learning.

2.6 Preventing Representation Collapse

As design recoverability requires the representation to distinguish different geometries, we additionally constrain variation across design cases. Standard SIGReg (Balestriero & LeCun, 2025) operates on the flattened token ensemble and thus does not guarantee diversity across cases. In our setting, representations can exhibit a near-constant token pattern across designs, with between-case standard deviation of approximately 0.0032, while satisfying the token-level SIGReg constraint. We address this failure mode with three complementary terms. First, case-level SIGReg is applied to the mean target representation

$$\mathcal{L}_{\text{sig-case}} = \text{SIGReg}\left(\frac{1}{M} \sum_m Z_{t,m} \right). \tag{9}$$

Second, we explicitly constrain the variance of the representation across design cases:

$$\mathcal{L}_{\text{var}} = \frac{1}{d} \sum_j \text{ReLU}\left(\gamma - \sqrt{\text{var}_{\text{case}}(\bar{Z}_t)_j + \epsilon} \right). \tag{10}$$

Finally, we require the target representation to retain sufficient information to reconstruct the physical field:

$$\mathcal{L}_{\text{rec-t}} = \mathbb{E}_q \left[\| D_t(Z_t, q) - F_s(q) \|_2^2 \right]. \tag{11}$$

We set $\lambda_c = \lambda_v = \lambda_t = 1$. Together, these constraints increase the observed between-case standard deviation to approximately 1.0 across the evaluated datasets, preventing the representation from satisfying token-level diversity constraints while remaining effectively constant across geometries.

2.7 Physics Decoder Transfer

After Stage 1, the geometry encoder is frozen and a new physics decoder $\phi_{s'}$ is trained for a new physical response space. This follows the parameter-efficient adaptation strategy used elsewhere in representation learning, where a frozen backbone is paired with a small trainable module (Houlsby et al., 2019; Hu et al., 2022). Given $g = E_\vartheta(P, N)$, where gradients are not propagated through E_ϑ, the transferred decoder is optimized according to

$$\min_{\phi_{s'}} \mathbb{E} \left[\left\| D_{\phi_{s'}}(g, q, c) - F_{s'}(q) \right\|_2^2 \right]. \tag{12}$$

Because the geometry representation is fixed, this experiment tests whether the latent captures geometric information that can be reused by a new physical decoder. Transfer efficiency therefore depends on whether g encodes reusable geometric features rather than task-specific information.

Table 1: Surrogate accuracy across parameterized 3D aerodynamic benchmarks. Full-field R^2 is reported for surface pressure and wall-shear-stress components, together with integrated force coefficients C_D and C_L. Values are mean $\pm$ std over three seeds; bold indicates the best value for each dataset and quantity.

Model	Dataset	Pressure	WSS_x	WSS_y	WSS_z	C_D	C_L
DoMINO	SHIFT-Wing	0.9678 ± 0.001	0.8972 ± 0.001	0.9456 ± 0.004	0.9654 ± 0.002	0.940 ± 0.002	0.891 ± 0.001
	SHIFT-Submarine	0.9321 ± 0.000	0.9332 ± 0.000	0.9421 ± 0.003	0.9243 ± 0.002	0.973 ± 0.001	–
	SHIFT-SUV	0.9765 ± 0.000	0.9475 ± 0.000	0.9241 ± 0.002	0.9342 ± 0.002	0.980 ± 0.003	–
GeoTransolver	SHIFT-Wing	0.9950 ± 0.001	0.9350 ± 0.001	0.9645 ± 0.002	0.9899 ± 0.001	0.960 ± 0.002	0.956 ± 0.001
	SHIFT-Submarine	0.9921 ± 0.001	0.9432 ± 0.001	0.9621 ± 0.001	0.9743 ± 0.002	0.990 ± 0.002	–
	SHIFT-SUV	0.9931 ± 0.001	0.9698 ± 0.001	0.9641 ± 0.001	0.9542 ± 0.001	0.996 ± 0.002	–
AeroJEPA	SHIFT-Wing	0.9892 ± 0.009	0.9168 ± 0.003	0.9618 ± 0.007	0.9719 ± 0.007	0.973 ± 0.011	0.952 ± 0.046
	SHIFT-Submarine	0.9937 ± 0.000	0.9532 ± 0.001	0.9602 ± 0.002	0.9764 ± 0.001	0.987 ± 0.003	–
	SHIFT-SUV	0.9870 ± 0.001	0.9352 ± 0.001	0.9522 ± 0.001	0.9437 ± 0.001	1.000 ± 0.000	–
D-JEPA	SHIFT-Wing	$\mathbf{0.9971 \pm 0.001}$	$\mathbf{0.9376 \pm 0.002}$	$\mathbf{0.9849 \pm 0.001}$	$\mathbf{0.9900 \pm 0.002}$	$\mathbf{0.990 \pm 0.002}$	$\mathbf{0.970 \pm 0.005}$
	SHIFT-Submarine	$\mathbf{0.9967 \pm 0.001}$	$\mathbf{0.9691 \pm 0.000}$	$\mathbf{0.9766 \pm 0.000}$	$\mathbf{0.9809 \pm 0.001}$	$\mathbf{0.995 \pm 0.002}$	–
	SHIFT-SUV	$\mathbf{0.9937 \pm 0.001}$	$\mathbf{0.9747 \pm 0.002}$	$\mathbf{0.9649 \pm 0.003}$	$\mathbf{0.9782 \pm 0.002}$	$\mathbf{1.000 \pm 0.000}$	–

3 Results & Discussion

We evaluate the learned geometry representation along three properties: (1) predictive accuracy for full-field and integrated quantities, (2) design information retention and reuse across operating conditions, and (3) transfer to new physical response spaces through swappable decoders. Finally, we evaluate differentiable design optimization with high-fidelity CFD validation.

3.1 Experimental datasets

We evaluate D-JEPA on four complementary 3D benchmarks. SHIFT-Wing (Luminary Cloud, 2025b) provides highly parameterized 3D aircraft wing geometries spanning transonic flow regimes, with varied aspect ratios, sweep and spanwise twist distributions. SHIFT-Submarine (Luminary Cloud, 2026) contains parameterized axisymmetric and non-axisymmetric submarine hulls featuring variations in fore-body shape, sail geometry and tail half-angles. SHIFT-SUV (Luminary Cloud, 2025a) covers parameterized automotive geometries that capture complex bluff-body aerodynamic phenomena, including slant angle, backlight geometry and underbody tapering. Finally, AutoHood 3D (Sharma et al., 2025) pairs automotive hood geometries with coupled fluid–structure interaction (FSI) data, in which aerodynamic surface pressure fields act as mechanical loads that generate structural deformation responses.

3.2 Full-field surrogate accuracy

We establish that D-JEPA retains strong full-field-prediction accuracy while introducing an explicitly structured geometry representation. Table 1 reports the coefficient of determination (R^2) for surface pressure, the three components of wall shear stress (WSS_x, WSS_y, WSS_z) and the integrated force coefficients (C_D, C_L), for D-JEPA and three baselines: the AeroJEPA predictive-latent model (Giral et al., 2026) and the field-reconstruction surrogates GeoTransolver (Adams et al., 2025) and DoMINO (Ranade et al., 2025). Implementation details for the baselines are given in Appendix F. Across all three aerodynamic benchmarks, D-JEPA maintains or improves field-prediction accuracy relative to AeroJEPA. Surface pressure prediction reaches $R^2 = 0.9971$ on SHIFT-Wing, 0.9967 on SHIFT-Submarine and 0.9937 on SHIFT-SUV. For wall shear stress, D-JEPA improves across every component, with WSS_y and WSS_z improving on all benchmarks. Relative to the full baseline set, D-JEPA attains the best value in all applicable dataset–channel combinations. D-JEPA also preserves accuracy for integrated aerodynamic observables. On SHIFT-Wing, C_D prediction improves from $R^2 = 0.973$ with AeroJEPA to 0.990, while C_L improves from 0.952 to 0.970. On SHIFT-SUV, D-JEPA matches the best C_D prediction at $R^2 = 1.000$. These results show that the structured representation preserves the predictive performance required for both spatially resolved fields and integrated engineering quantities.

Table 2: Case-level variation in the target representation under progressively stronger representation constraints. Values report $\mathrm{std}_{\mathrm{case}}(\mathrm{mean}_m Z_t)$, measuring variation across geometry cases.

Dataset	None	Token SIGReg	Case SIGReg	+ Variance	+ Target Recon
SHIFT-Wing	0.0003	0.0003	0.6211	1.0346	1.0317
SHIFT-Submarine	0.0004	0.0005	0.7566	1.0633	1.0620
SHIFT-SUV	0.0002	0.0002	0.0003	0.9892	0.9889

3.3 Geometry representation and transfer capabilities

Having established full-field surrogate accuracy, we evaluate the properties that distinguish D-JEPA from a conventional predictive surrogate: case-level representation diversity, design-variable recoverability, reuse across operating conditions and cross-physics decoder transfer.

3.3.1 Preventing case-level representation collapse

Design recoverability requires the latent to distinguish different geometries, but token-level diversity alone does not guarantee variation across design cases. We therefore evaluate the effect of the representation constraints on between-case variation. Table 2 summarizes this ablation. Without case-level constraints, the between-case standard deviation of the target representation remains near zero (0.0002–0.0004) across all three benchmarks. Adding token-level SIGReg alone produces essentially no change, showing that token-level diversity does not prevent case-level collapse. Case-level constraints substantially increase between-case variation. The variance hinge and target-reconstruction terms increase the standard deviation to approximately 1.0 across all three geometry families, reaching 1.03–1.06 on SHIFT-Wing and SHIFT-Submarine and approximately 0.99 on SHIFT-SUV. Importantly, these changes occur with only modest changes in forward pressure prediction, indicating that the additional constraints primarily alter the structure of the representation rather than its predictive capacity. The complete diagnostics, including latent and reconstruction losses, are reported in Appendix D.

3.3.2 Design-variable recoverability

A central objective of D-JEPA is to retain information about the underlying geometric design variables in the learned latent. We therefore evaluate whether the design parameters can be recovered using a linear probe from g. Table A.5 in Appendix G reports recovery R^2 for every design variable of SHIFT-SUV, SHIFT-Wing and SHIFT-Submarine. D-JEPA achieves near-perfect recoverability across the parameterized shape dimensions, with mean R^2 scores of 0.989 on SHIFT-SUV, 0.998 on SHIFT-Wing and 0.998 on SHIFT-Submarine. As a representation-level reference, we apply the identical probe to the AeroJEPA context latent, mean-pooled to the same 128 dimensions as g. The corresponding mean R^2 values are 0.683, 0.801 and 0.882, respectively, giving gains of 0.306, 0.197 and 0.116. Thus, D-JEPA retains substantially more linearly recoverable design information across all three geometry families. This structured decodability provides a direct representation-level basis for design navigation and complements the full-field prediction results: the latent is not only useful for predicting physical fields, but also retains explicit information about the geometric variables defining the input shapes. We further perform component ablations to isolate the sources of this recoverability. Here *Learned-only* denotes the point-transformer branch alone, with the parameter-free global and regional descriptors removed, so that the comparison isolates the contribution of the coordinate-aware descriptors. As shown in Table 3, removing the design loss $\mathcal{L}_{\mathrm{dis}}$ substantially reduces design recovery across all three datasets, while having a comparatively smaller effect on pressure prediction. Removing the latent prediction loss $\mathcal{L}_{\mathrm{lat}}$ produces a smaller but consistent reduction in design and pressure R^2 across the three datasets. The learned-only encoder retains substantial design information, indicating that recoverability is not solely due to the explicit geometric descriptors.

3.3.3 Operating-condition sweeps (same physics, new conditions)

As the geometry encoder E_ϑ receives neither the condition vector c nor the physics label s, the latent representation g can be computed once per geometry and reused across operating conditions. We evaluate this capability on SHIFT-Wing across a transonic Mach sweep ($\mathrm{Mach} \in [0.50, 0.90]$).

Table 3: Component ablation of D-JEPA across the SHIFT datasets, showing the contributions of latent prediction and design supervision to design recoverability and pressure prediction. *Learned-only* removes the parameter free global and regional descriptors.

	Design R^2 ↑				Pressure R^2 ↑			
Dataset	**Full**	$-\mathcal{L}_{\text{lat}}$	**Learned-only**	$-\mathcal{L}_{\text{dis}}$	**Full**	$-\mathcal{L}_{\text{lat}}$	**Learned-only**	$-\mathcal{L}_{\text{dis}}$
SHIFT-Wing	**0.998**	0.993	0.976	0.722	**0.9971**	0.9912	0.9885	0.9823
SHIFT-Submarine	**0.998**	0.982	0.757	0.564	**0.9967**	0.9896	0.9852	0.9872
SHIFT-SUV	**0.989**	0.984	0.976	0.622	**0.9937**	0.9851	0.9801	0.9821

Table 4: SHIFT-Wing pressure prediction accuracy (R^2) across trained and held-out Mach numbers using a single frozen geometry representation g. The held-out Mach values are interpolation points within the trained Mach range. Bold values denote held-out conditions.

	Trained operating conditions										**Held-out**	
Mach	0.5000	0.6000	0.7000	0.7500	0.8000	0.8250	0.8375	0.8500	0.8750	0.9000	**0.7750**	**0.8125**
Pressure R^2	0.9902	0.9935	0.9948	0.9944	0.9941	0.9956	0.9960	0.9967	0.9952	0.9939	**0.9922**	**0.9944**
Std	0.001	0.002	0.001	0.003	0.002	0.001	0.001	0.002	0.001	0.002	0.001	0.002

Ten Mach values are used during training, while Mach $= 0.7750$ and Mach $= 0.8125$ are held out entirely. Importantly, both held-out values lie within the trained Mach range and therefore test *interpolation* across operating conditions rather than extrapolation. Table 4 reports pressure prediction R^2 across all trained and held-out conditions. The frozen geometry representation maintains $R^2 \geq 0.9902$ across all evaluated conditions, including $R^2 = 0.9922$ and 0.9944 at the two held-out Mach numbers. A single geometry representation therefore remains reusable across the continuous operating sweep without additional geometry-encoder training.

3.3.4 Swappable physics decoders (cross-physics transfer)

We next test whether the same geometry representation can be reused when the target physical response changes. We evaluate this capability on AutoHood 3D by transferring a geometry encoder pretrained on aerodynamic pressure fields to structural displacement prediction. In the D-JEPA transfer setting, the pretrained aerodynamic geometry encoder is completely frozen and only a lightweight structural FiLM decoder is trained, requiring 0.56 M trainable parameters. In contrast, the from-scratch baseline trains the full geometry encoder and structural decoder end-to-end, requiring 2.30 M trainable parameters. Table 5 tracks structural prediction R^2 across training budgets. The swappable-decoder formulation reduces trainable parameters by $4.1\times$ and provides substantially faster adaptation at low training budgets. After 10 epochs, the transferred representation reaches $R^2 = 0.3316$, compared with 0.0396 from scratch. At 100 epochs, it reaches 0.8901 versus 0.8740, and the transferred model remains ahead through 300 epochs, by which point the two approaches are within 0.007 of each other. The benefit demonstrated here is therefore parameter-efficient and rapid adaptation: the aerodynamic geometry representation can be reused for this CFD-to-structural transfer while training only a lightweight physics-specific decoder.

3.4 Design optimization & CFD validation

Finally, we evaluate the representation in a downstream engineering workflow using gradient-based design optimization, with selected designs validated against high-fidelity CFD. The differentiable optimization path and design-space constraints are detailed in Appendix H. To assess whether surrogate-guided optimization produces physically consistent design rankings, we perform direct high-fidelity CFD validation for SHIFT-Submarine and SHIFT-Wing. Equivalent high-fidelity CFD validation for SHIFT-SUV was not possible because the corresponding high-fidelity CFD setup and sufficient geometry and operating-condition information were not available. For each benchmark, we evaluate four candidates selected from the optimization trajectories. *OPT-*τ is the predicted best design within the nominal design radius τ, representing a "safe" optimum. *OPT-Wide* is the predicted best design with the radius relaxed to 1.8τ, explicitly testing extrapolation beyond the nominal region. *ANTI* is the predicted worst design within τ and serves as a negative control, while *MID* is a held-out design with a mid-range predicted objective and provides a buildable sanity check.

Table 5: Structural displacement prediction accuracy (R^2) on AutoHood 3D comparing frozen geometry encoder transfer versus training from scratch.

Training epoch budget	Frozen geometry encoder (0.56 M params)	From scratch (2.30 M params)
10	**0.3316 ± 0.020**	0.0396 ± 0.007
100	**0.8901 ± 0.013**	0.8740 ± 0.034
200	**0.9327 ± 0.002**	0.9297 ± 0.012
300	**0.9385 ± 0.001**	0.9313 ± 0.002

For SHIFT-Submarine, CFD is evaluated at $\rho = 998.2\,\mathrm{kg/m^3}$, $V = 5.0\,\mathrm{m/s}$, $q = 12{,}478\,\mathrm{Pa}$ and $A_{\mathrm{ref}} = 19.27\,\mathrm{m^2}$. Predicted drag agrees with CFD within 0.6–2.0%. For SHIFT-Wing, the relative error in integrated lift ranges from 0.72–10.74%. Despite these differences in absolute error, the surrogate exactly preserves the CFD ranking of all four candidates in both benchmarks, with rankings $1 \to 1$, $2 \to 2$, $3 \to 3$ and $4 \to 4$, as reported in Table 6. These results provide a direct high-fidelity check that the surrogate preserves the relative ordering of optimized and control designs, including the extrapolative *OPT-Wide* case.

Table 6: High-fidelity CFD validation of surrogate-optimized designs. Integrated drag is reported for SHIFT-Submarine and integrated lift for SHIFT-Wing. Predicted and CFD rankings agree for all evaluated designs.

Benchmark	Design candidate	Predicted [N]	CFD [N]	Relative error	Predicted → CFD rank
SHIFT-Submarine (drag)	OPT-wide	645.8	637.69	1.3%	**1 → 1**
	OPT-tau	677.0	688.36	1.7%	**2 → 2**
	MID	718.6	723.08	0.6%	**3 → 3**
	ANTI	758.7	773.75	2.0%	**4 → 4**
SHIFT-Wing (lift)	OPT-Wide	3,426,080	3,193,866	7.27%	**1 → 1**
	OPT-Tau	3,108,536	3,086,237	0.72%	**2 → 2**
	MID	2,008,248	2,249,932	10.74%	**3 → 3**
	ANTI	1,434,578	1,386,571	3.46%	**4 → 4**

The experiments show that D-JEPA provides a reusable geometry representation for accurate simulation, transfer and design optimization, improving full-field prediction over the evaluated field-reconstruction and predictive-latent baselines while maintaining strong integral accuracy and enabling reuse across design spaces, operating conditions and downstream physics without retraining the geometry encoder. Design-variable recovery further provides a differentiable path from the physical objective back to the design parameters, which supports gradient-based design exploration. However, the evaluation remains limited to a small set of aerodynamic and structural benchmarks, with physics transfer demonstrated only in the CFD-to-structural setting and optimization restricted to differentiable surrogates with predefined objectives; broader multi-physics validation, diverse geometries and constrained real-world design problems remain important directions for future work.

4 Conclusions

We present D-JEPA, a geometry-centric representation learning framework designed to separate reusable geometric information from physics- and condition-specific prediction. Across three challenging parameterized three-dimensional aerodynamic benchmarks, D-JEPA maintains or improves full-field surrogate accuracy relative to the evaluated baselines while maintaining high accuracy for integral aerodynamic quantities. We show that the learned geometric representation can be reused across different geometry families and operating conditions, and transferred to a structural response task on AutoHood 3D without retraining the geometry encoder. Finally, the learned representation supports gradient-based design optimization, with the resulting candidates validated against high-fidelity CFD. Together, these results demonstrate that explicitly separating geometry representation from physics-specific prediction can provide a reusable geometric representation by design, enabling a single learned representation to support multiple prediction tasks and downstream optimization.

Ethics Statement

This work does not involve human subjects, personally identifiable information or sensitive personal data. The datasets used in this study consist of scientific simulation and engineering data. To the best of our knowledge, the proposed methodology does not introduce additional privacy, security or safety risks beyond those inherent to the underlying scientific and engineering applications.

Reproducibility Statement

We provide detailed descriptions of the D-JEPA architecture, training procedure, datasets, preprocessing, evaluation protocols and experimental settings in the main paper and supplementary material. The appendix further documents the model architecture, hyperparameters, training configuration, dataset construction and evaluation metrics used in our experiments. These details are provided to facilitate independent reproduction of the reported results. Where permitted by dataset and project constraints, source code and experiment configuration files will be provided in the supplementary material.

AI Use Statement

We used large language model (LLM) tools to obtain feedback on the manuscript, primarily to identify opportunities for improving clarity, organization and presentation. The feedback was used in accordance with the ICLR author guidelines to improve the manuscript. LLMs were not used to develop the scientific methodology, formulate hypotheses, design or conduct experiments, analyze or interpret scientific results, generate scientific claims or produce the technical contributions presented in this work. The authors independently conducted and verified all scientific work and are fully responsible for the content, claims and conclusions of this paper.

References

Corey Adams, Rishikesh Ranade, Ram Cherukuri, and Sanjay Choudhry. GeoTransolver: Learning physics on irregular domains using multi-scale geometry aware physics attention transformer. *arXiv preprint arXiv:2512.20399*, 2025.

Guillaume Alain and Yoshua Bengio. Understanding intermediate layers using linear classifier probes. In *International Conference on Learning Representations (ICLR) Workshop*, 2017.

Benedikt Alkin, Maurits Bleeker, Richard Kurle, Tobias Kronlachner, Reinhard Sonnleitner, Matthias Dorfer, and Johannes Brandstetter. AB-UPT: Scaling neural CFD surrogates for high-fidelity automotive aerodynamics simulations via anchored-branched universal physics transformers. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=nwQ8nitlTZ.

Mathilde aron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9650–9660, 2021.

Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15619–15629, 2023.

Alexei Baevski, Wei-Ning Hsu, Qiantong Xu, Arun Babu, Jiatao Gu, and Michael Auli. data2vec: A general framework for self-supervised learning in speech, vision and language. In *International Conference on Machine Learning (ICML)*, volume 162, pp. 1298–1312, 2022.

Randall Balestriero and Yann LeCun. LeJEPA: Provable and scalable self-supervised learning without the heuristics. *arXiv preprint arXiv:2511.08544*, 2025.

Adrien Bardes, Jean Ponce, and Yann LeCun. VICReg: Variance-invariance-covariance regularization for self-supervised learning. In *International Conference on Learning Representations (ICLR)*, 2022.

Adrien Bardes, Quentin Garrido, Jean Ponce, Xinlei Chen, Michael Rabbat, Yann LeCun, Mahmoud Assran, and Nicolas Ballas. V-JEPA: Revisiting feature prediction for learning visual representations from video. *arXiv preprint arXiv:2404.08471*, 2024.

Steven L. Brunton, Bernd R. Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics*, 52:477–508, 2020.

M. A. Frederick et al. Development of the X-57 aerodynamic database CFD results. Technical report, NASA, 2025. URL https://ntrs.nasa.gov/api/citations/20250001715/downloads/20250001715%20FINAL.pdf.

Francisco Giral, Abhijeet Vishwasrao, Andrea Arroyo Ramo, Mahmoud Golestanian, Federica Tonti, Adrian Lozano-Durán, Steven L. Brunton, Sergio Hoyas, Hector Gomez, Soledad Le Clainche, and Ricardo Vinuesa. AeroJEPA: Learning semantic latent representations for scalable 3D aerodynamic field modeling. *arXiv preprint arXiv:2605.05586*, 2026.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. Bootstrap your own latent: A new approach to self-supervised learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, 2020.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin de Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for NLP. In *International Conference on Machine Learning (ICML)*, 2019.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations (ICLR)*, 2022.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2019.

Luminary Cloud. SHIFT-SUV: High-fidelity computational fluid dynamics dataset for SUV external aerodynamics, 2025a. URL https://huggingface.co/datasets/luminary-shift/SUV/.

Luminary Cloud. SHIFT-Wing: High-fidelity computational fluid dynamics dataset for transonic aerospace external aerodynamics, 2025b. URL https://huggingface.co/datasets/luminary-shift/Wing.

Luminary Cloud. SHIFT-Submarine: High-fidelity computational fluid dynamics dataset for submarine hydrodynamics, 2026. URL https://huggingface.co/datasets/luminary-shift/Submarine.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

NASA Advanced Supercomputing Division. Building aerodynamic databases for the SLS design process. NASA Advanced Supercomputing Division, 2017. URL https://www.nas.nasa.gov/SC17/demos/demo11.html.

NASA Advanced Supercomputing Division. Aerodynamic databases for the Space Launch System. NASA Advanced Supercomputing Division, 2018. URL https://ntrs.nasa.gov/api/citations/20220016689/downloads/Meeroff_SC22_presentation_lowres.pdf.

NASA Advanced Supercomputing Division. CFD aerodynamic database for NASA X-57 Maxwell airplane. NASA Advanced Supercomputing Division, 2021. URL https://www.nas.nasa.gov/SC21/research/project22.html.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual reasoning with a general conditioning layer. In *AAAI Conference on Artificial Intelligence*, 2018.

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J. Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Rishikesh Ranade, Mohammad A. Nabian, Akshay Tangsali, Alexey Kamenev, Oliver Hennigh, Ram Cherukuri, and Sanjay Choudhry. DoMINO: A decomposable multi-scale iterative neural operator for modeling large scale engineering simulations. *arXiv preprint arXiv:2501.13350*, 2025.

Vansh Sharma, Harish Jai Ganesh, Maryam Akram, Wanjiao Liu, and Venkat Raman. AutoHood3D: A multi-modal benchmark for automotive hood design and fluid-structure interaction, 2025. URL https://arxiv.org/abs/2511.05596.

Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for PDEs on general geometries. In *International Conference on Machine Learning (ICML)*, 2024a.

Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, and Hengshuang Zhao. Point transformer V3: Simpler, faster, stronger. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4840–4851, 2024b.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow Twins: Self-supervised learning via redundancy reduction. In *International Conference on Machine Learning (ICML)*, volume 139, pp. 12310–12320, 2021.

APPENDIX

This appendix provides additional details on dataset construction and preprocessing, model architecture, training objectives, representation collapse diagnostics, evaluation protocols, baseline implementations, design-variable recovery, operating-condition generalization, cross-physics transfer, differentiable optimization and computational requirements. It is intended to provide sufficient implementation and evaluation detail to reproduce the experiments reported in the main paper.

A DATASET SPECIFICATIONS AND PREPROCESSING PIPELINES

We evaluate D-JEPA across four benchmark datasets spanning parameterized three-dimensional physical systems: SHIFT-Wing (Luminary Cloud, 2025b), SHIFT-Submarine (Luminary Cloud, 2026), SHIFT-SUV (Luminary Cloud, 2025a) and AutoHood 3D (Sharma et al., 2025). Each case is represented by an unstructured spatial point cloud P with associated outward surface normals N, solver-generated physical fields, underlying geometric design parameters and external operating conditions when applicable. Table A.1 summarizes the benchmark characteristics.

Table A.1: Benchmark dataset properties, parameter dimensions, operating conditions and surface targets. A dash indicates that the quantity is not applicable.

Dataset	Cases	Design parameters	Operating conditions	Surface fields
SHIFT-Wing	1,966	7	angle of attack, Mach	pressure, WSS_x, WSS_y, WSS_z
SHIFT-Submarine	4,000	7	fixed	pressure, WSS_x, WSS_y, WSS_z
SHIFT-SUV	1,996	7	fixed	pressure, WSS_x, WSS_y, WSS_z
AutoHood 3D	12,000	–	load case	pressure, displacement

SHIFT-Wing is parameterized by seven geometric variables comprising aspect ratio, quarter-chord sweep angle $\Lambda_{c/4}$ and five spanwise twist variables. SHIFT-Submarine contains seven parameters controlling cross-sectional geometry, fore-body shape and length, mid-body length, aft-tail half-angle, sail-length ratio and strake-length ratio. SHIFT-SUV contains seven parameters describing hood height, hood-base height, windshield inclination, total vehicle height, backlight angle, rear tapering ratio and front-rear planform curvature.

The explicit design parameters are not provided to the physics decoder. Instead, they are used only by the auxiliary design-recoverability objective during training. At inference, the decoder receives the learned geometry representation g, query location q, local surface normal $n(q)$ and operating condition vector c. This separation allows the design variables to be evaluated as information retained by the geometry representation rather than as direct decoder inputs.

AutoHood 3D provides a cross-physics evaluation in which the same geometric representation can support different physical response spaces. The geometry is associated with an aerodynamic pressure field and a structural displacement response. D-JEPA therefore uses the same geometry representation while allowing physics-specific decoders to predict the corresponding outputs. For SHIFT-Submarine and SHIFT-SUV, the operating condition vector is empty because the evaluated flow conditions are fixed. Operating-condition conditioning is explicitly evaluated on SHIFT-Wing through variation in Mach number and angle of attack.

B DETAILED MODEL ARCHITECTURE

B.1 GEOMETRY ENCODER

The geometry encoder E_ϑ maps an unstructured point cloud and its surface normals to a compact geometry representation, as given in Eq. (A.1):

$$g = E_\vartheta(P, N) \in \mathbb{R}^{128}. \tag{A.1}$$

The encoder first processes the input geometry into $M = 512$ context tokens with feature dimension $d = 128$. Local spatial neighborhoods use $k_{\mathrm{aggregate}} = 16$ nearest neighbors for message passing

and $k_{\text{attn}} = 16$ neighbors for local self-attention. The backbone contains six transformer blocks with eight attention heads and an MLP expansion factor of four. The resulting token matrix $Z \in \mathbb{R}^{512\times128}$ is summarized using channel-wise mean and maximum pooling, as given in Eq. (A.2):

$$s_{\text{loc}} = \left[\operatorname*{mean}_m Z_m \,;\, \max_m Z_m\right] \in \mathbb{R}^{256}. \tag{A.2}$$

The local representation is augmented with non-parametric global and regional descriptors to provide explicit information about large-scale geometry.

B.2 Global and regional geometric descriptors

The global descriptor s_{glob} contains 15 geometric statistics computed directly from the spatial coordinates. These include per-axis coordinate means and standard deviations, oriented bounding-box dimensions, eigenvalues of the spatial covariance matrix and off-diagonal covariance terms.

The regional descriptor s_{reg} discretizes the canonical bounding domain into a $5 \times 5 \times 5$ voxel grid. For each voxel, local coordinate statistics, surface-normal statistics and point-density information are computed. The resulting descriptor contains $125 \times 13 = 1{,}625$ features.

The three representations are concatenated as in Eq. (A.3):

$$s_{\text{fused}} = [s_{\text{loc}} \,;\, s_{\text{glob}} \,;\, s_{\text{reg}}] \in \mathbb{R}^{1896}, \tag{A.3}$$

and passed through an MLP to produce the final 128-dimensional geometry representation g.

Neither the operating condition vector c nor the physics identifier s is provided to the geometry encoder. Thus, the geometry representation is computed solely from geometric information and can subsequently be reused when the operating condition or physical response space changes.

B.3 JEPA representation-learning pathway

D-JEPA uses latent prediction as an auxiliary representation-learning objective. During training, a context representation is used to predict target latent tokens through the JEPA predictor. The resulting latent prediction objective encourages the geometry representation to retain information useful for predicting the target representation. The JEPA predictor is not used as the physics-field decoder. Instead, the learned geometry representation g is passed directly to the physics-specific decoder for downstream field prediction. This separation allows the JEPA pathway to shape the geometry representation while keeping the physical output mapping independently swappable.

B.4 FiLM-modulated physics decoder

For a physical response space s, the decoder is written as in Eq. (A.4):

$$\hat{F}_s(q) = D_{\phi_s}(g, q, c), \tag{A.4}$$

where g is the geometry representation, q is the query location, $n(q)$ is the local surface normal and c denotes the operating condition. Query coordinates are encoded using Fourier features across 16 frequency octaves (Tancik et al., 2020), producing a 99-dimensional positional representation. The decoder contains four residual FiLM blocks with hidden width 256. The geometry representation modulates the query pathway through learned affine scale and shift parameters rather than being treated only as a concatenated input (Perez et al., 2018). This factorization separates geometry representation learning from the mapping from geometry and operating conditions to a particular physical response space. Consequently, a new decoder can be trained while keeping the geometry encoder fixed. For the aerodynamic benchmarks, the decoder predicts four surface quantities: pressure and the three components of wall shear stress (WSS_x, WSS_y and WSS_z).

C Training Objectives and Optimization

C.1 Composite training objective

D-JEPA is trained end-to-end using a composite objective that combines latent prediction, physical reconstruction, design-variable recovery, representation regularization and target reconstruction, as

given in Eq. (A.5):

$$\mathcal{L} = \mathcal{L}_{\text{lat}} + \lambda_r \mathcal{L}_{\text{rec}} + \lambda_d \mathcal{L}_{\text{dis}} + \lambda_s \mathcal{L}_{\text{sig}} + \lambda_c \mathcal{L}_{\text{sig-case}} + \lambda_v \mathcal{L}_{\text{var}} + \lambda_t \mathcal{L}_{\text{rec-t}}. \tag{A.5}$$

The latent prediction term $\mathcal{L}_{\text{lat}}$ aligns predicted latent representations with target-encoder representations. The physical reconstruction loss $\mathcal{L}_{\text{rec}}$ trains the physics decoder to predict the target physical field. The design-recoverability loss $\mathcal{L}_{\text{dis}}$ trains a linear probe to recover the known design variables from the geometry representation. The token-level SIGReg term $\mathcal{L}_{\text{sig}}$ (Balestriero & LeCun, 2025) encourages diversity within the target-token representation. Because token-level diversity does not by itself guarantee variation between different geometry cases, D-JEPA also evaluates case-level regularization. The case-level SIGReg term $\mathcal{L}_{\text{sig-case}}$ and variance constraint $\mathcal{L}_{\text{var}}$ encourage variation across geometry cases, while the target-reconstruction term $\mathcal{L}_{\text{rec-t}}$ provides an additional connection between target representations and local physical coordinates. The resulting objective is designed to jointly maintain predictive accuracy, design information and non-collapsed geometry representations. The corresponding weights are listed in Table A.2.

Table A.2: Hyperparameter weights used for the composite D-JEPA objective.

Loss term	Latent ($\mathcal{L}_{\text{lat}}$)	Recon. ($\lambda_r \mathcal{L}_{\text{rec}}$)	Design ($\lambda_d \mathcal{L}_{\text{dis}}$)	Token ($\lambda_s \mathcal{L}_{\text{sig}}$)	Case ($\lambda_c \mathcal{L}_{\text{sig-case}}$)	Var. ($\lambda_v \mathcal{L}_{\text{var}}$)	Tgt. rec. ($\lambda_t \mathcal{L}_{\text{rec-t}}$)
Weight	1.0	1.0	1.0	0.01	1.0	1.0	1.0

C.2 Optimization schedule

AeroJEPA and D-JEPA use matched optimization settings unless otherwise specified. The primary training configuration is summarized in Table A.3. The batch size is 16 for SHIFT-Wing, SHIFT-SUV and SHIFT-Submarine, selected to accommodate the memory requirements of the underlying solver meshes.

Table A.3: Optimization configuration and training hyperparameters.

Setting	**Value**
Optimizer	AdamW (Loshchilov & Hutter, 2019)
Base LR	1×10^{-3}
Weight decay	1×10^{-3}
LR schedule	Cosine decay
Gradient clipping	10.0
Epochs	400
Precision	FP32
Seeds	42, 43, 44

D Representation-Collapse Diagnostics

A central requirement of the geometry representation is that different geometry cases should provide sufficiently distinct learning signals to avoid case-level collapse. We therefore measure between-case variation in the target representation using Eq. (A.6):

$$\sigma_{\text{case}} = \text{std}_{\text{case}} \left(\underset{m}{\text{mean}} \, Z_t \right), \tag{A.6}$$

where Z_t denotes the target token representation and the mean is taken over tokens before computing the standard deviation across geometry cases. This diagnostic distinguishes case-level variation from token-level diversity: a target representation may contain diverse tokens within an individual geometry while remaining nearly identical across different geometries.

Table A.4 extends the summary in Table 2 with the latent and reconstruction losses and the corresponding downstream metrics. Two observations are relevant beyond the between-case standard

deviation itself. First, the latent loss L_{lat} is smallest in exactly the configurations where σ_{case} is near zero, so it cannot be used on its own as a training diagnostic. Second, the downstream metrics vary only modestly across configurations, because L_{rec} and the design probe act on the geometry representation g while the collapse is measured in the target tokens Z_t. Case-level variation therefore has to be monitored directly.

Table A.4: Complete representation-collapse diagnostics. The case-level standard deviation measures between-case variation, while L_{lat} and L_{rec} report latent prediction and physical reconstruction losses. Pressure R^2 and design-recovery R^2 quantify downstream performance.

Dataset	**Configuration**	σ_{case}	L_{lat}	L_{rec}	Pressure R^2	Design R^2
SHIFT-Wing	None	0.0003	0.0000	0.0354	0.9906	0.992
	Token SIGReg	0.0003	0.0000	0.0352	0.9906	0.992
	Case SIGReg	0.6211	0.0403	0.0373	0.9893	0.987
	+ Variance	1.0346	0.0498	0.0381	0.9886	0.986
	+ Target Recon.	1.0317	0.0447	0.0379	0.9971	0.998
SHIFT-Submarine	None	0.0004	0.0000	0.0348	0.9896	0.982
	Token SIGReg	0.0005	0.0000	0.0352	0.9897	0.982
	Case SIGReg	0.7566	0.0221	0.0361	0.9892	0.980
	+ Variance	1.0633	0.0221	0.0360	0.9894	0.982
	+ Target Recon.	1.0620	0.0192	0.0362	0.9967	0.998
SHIFT-SUV	None	0.0002	0.0000	0.0537	0.9726	0.975
	Token SIGReg	0.0002	0.0000	0.0534	0.9729	0.979
	Case SIGReg	0.0003	0.0000	0.0529	0.9734	0.982
	+ Variance	0.9892	0.1493	0.0547	0.9718	0.967
	+ Target Recon.	0.9889	0.1526	0.0551	0.9937	0.989

These results motivate the use of case-level constraints in the final D-JEPA training objective. They also illustrate why token-level diversity alone is insufficient as a diagnostic for a geometry representation: variation must be evaluated across geometry cases.

E Evaluation Protocols

All benchmark evaluations use fixed 10% validation partitions generated with constant random seeding. Validation metrics are evaluated on held-out geometry cases that are not used for model training. Field prediction accuracy is quantified using the coefficient of determination R^2, computed independently for each physical output variable and averaged across held-out cases. Reported uncertainty values are standard deviations across three independent training seeds. Integrated force coefficients, including drag C_D and lift C_L, are computed by direct integration of predicted pressure and wall-shear-stress fields over the complete solver surface. No secondary regression head is used to predict the integrated quantities. Design-variable recoverability is evaluated using a standardized, five-fold cross-validated ridge-regression probe applied to frozen 128-dimensional geometry representations g (Alain & Bengio, 2017). Out-of-fold R^2 values are reported to avoid evaluating the probe on the same samples used for fitting. Qualitative field comparisons are performed on dense, high-resolution solver meshes rather than only on the sparse point samples used during training. Representative median and worst-case held-out examples are shown in Figures A.1, A.2 and A.3.

F Baseline Models and Comparative Benchmarks

F.1 AeroJEPA baseline

AeroJEPA (Giral et al., 2026) is used as the primary representation-learning baseline. The comparison retains the same geometry backbone components, point-sampling procedure, optimizer settings and primary reconstruction objective where applicable. This controls for differences in model ca-

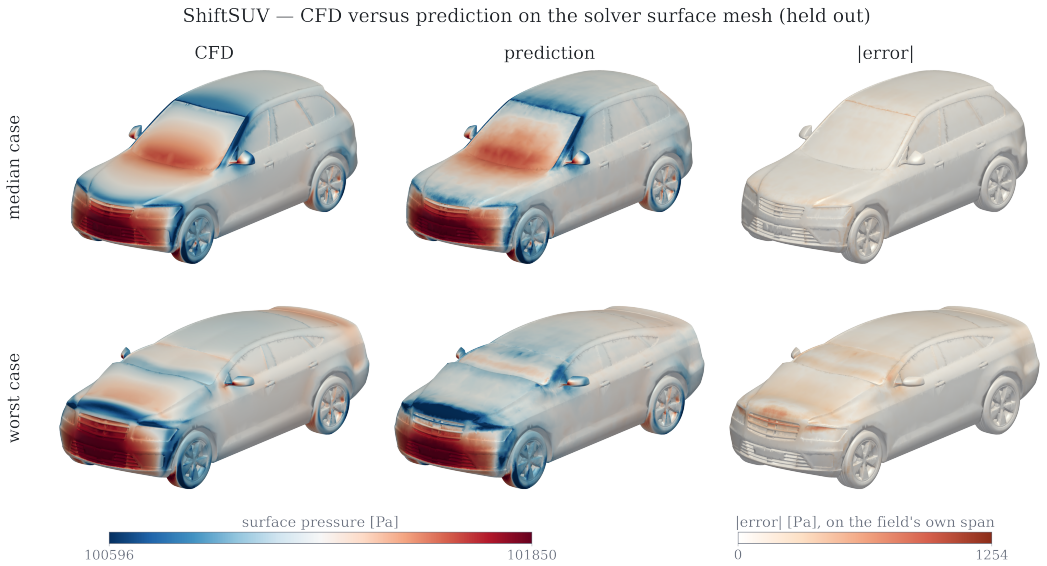


Figure A.1: Qualitative surface pressure prediction for SHIFT-SUV. Predicted pressure and absolute point-wise error are shown for representative median and worst-case held-out geometries.

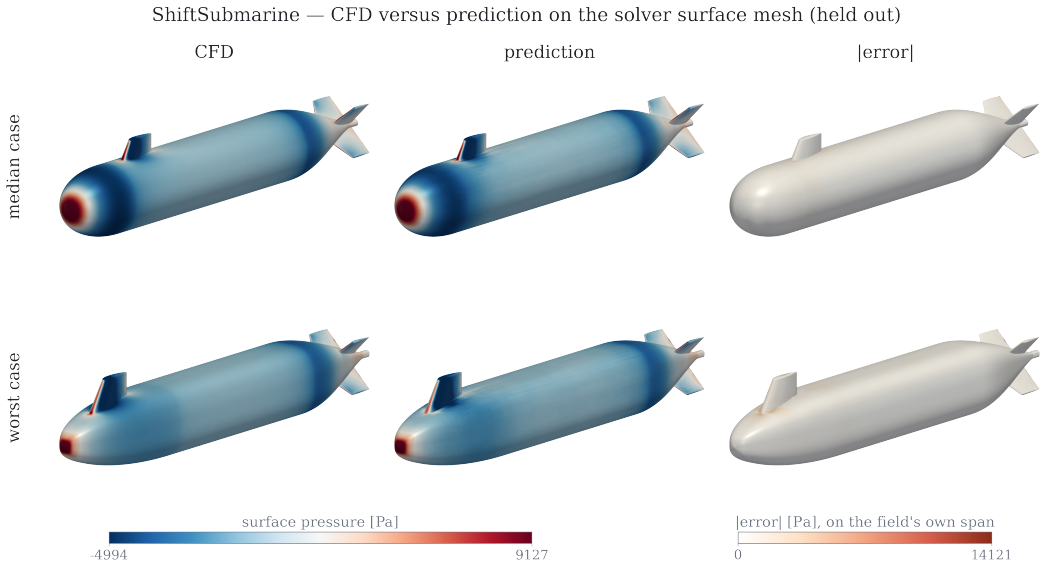


Figure A.2: Qualitative surface pressure prediction for SHIFT-Submarine. Predicted pressure and absolute point-wise error are shown for representative median and worst-case held-out geometries.

pacity and focuses the comparison on the geometry representation and additional D-JEPA training objectives. For design recoverability comparison, the pooled 128-dimensional AeroJEPA context representation is evaluated using the same linear probing procedure applied to D-JEPA.

F.2 GeoTransolver and DoMINO baselines

GeoTransolver (Adams et al., 2025) and DoMINO (Ranade et al., 2025) are evaluated using their official implementations on the same dataset splits. GeoTransolver uses 20 transformer layers, hidden width 256, eight attention heads, 128 slices and a 60,000-point surface representation under the corresponding optimization settings.

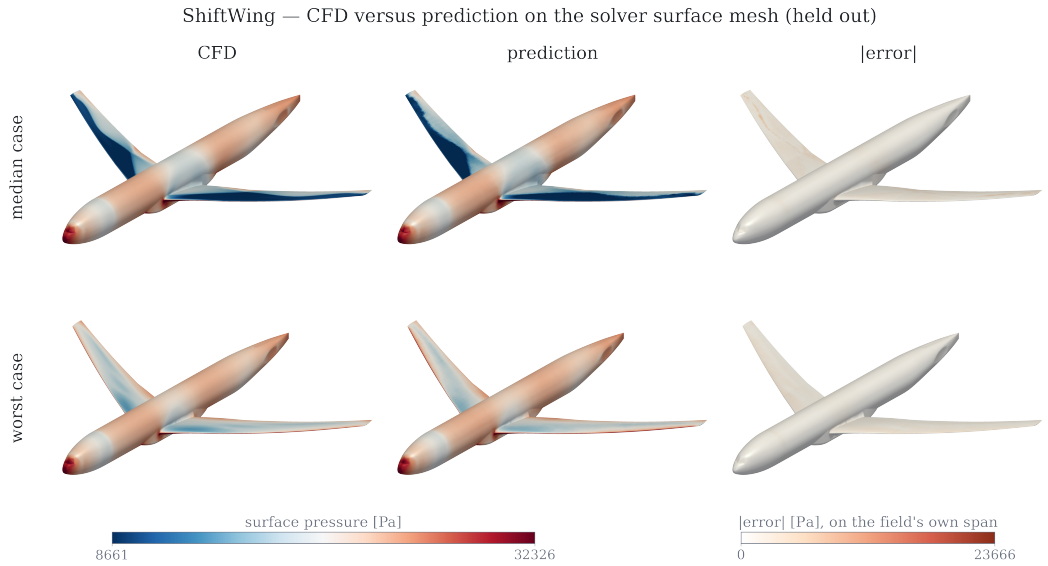


Figure A.3: Qualitative surface pressure prediction for SHIFT-Wing. Predicted pressure and absolute point-wise error are shown for representative median and worst-case held-out geometries.

All baseline metrics are computed on held-out validation cases. The resulting field-prediction and integrated-force comparisons are reported in Table 1 of the main text.

G Supplemental Experimental Results

G.1 Design-variable recoverability

A central property of D-JEPA is that the geometry representation retains information about the underlying design variables. We evaluate this property independently of the physics decoder by freezing the learned 128-dimensional geometry representation and fitting a linear ridge probe. Table A.5 reports the recovery accuracy for every design variable. Mean R^2 values are 0.989 for SHIFT-SUV, 0.998 for SHIFT-Wing and 0.998 for SHIFT-Submarine.

Table A.5: Linear-probe recovery accuracy (R^2) for individual geometric design parameters from the frozen 128-dimensional geometry representation g.

SHIFT-SUV		**SHIFT-Wing**		**SHIFT-Submarine**	
Parameter	R^2	Parameter	R^2	Parameter	R^2
hood_height	0.995	AR (aspect ratio)	0.999	cross_section_shape	1.000
hood_base_height	0.987	Lambda_c_4 (quarter-chord sweep)	0.998	fore_body_shape	0.996
windshield_angle	0.986	twist_20	0.997	fore_length	0.999
vehicle_height	0.991	twist_k	0.997	mid_length	0.999
backlight_angle	0.990	twist_60	0.999	aft_tail_half_angle	0.997
rr_tapering	0.988	twist_80	0.999	sail_length_ratio	0.997
fr_plainview	0.986	twist_t	0.998	strake_length_ratio	0.996
Mean	**0.989**	**Mean**	**0.998**	**Mean**	**0.998**

Individual design-parameter R^2 values range from 0.986 to 0.995 for SHIFT-SUV, 0.997 to 0.999 for SHIFT-Wing and 0.996 to 1.000 for SHIFT-Submarine. These results indicate that the learned geometry representation retains substantial linearly recoverable information about the underlying parameterization. Figure A.4 shows the corresponding per-parameter probe correlations.

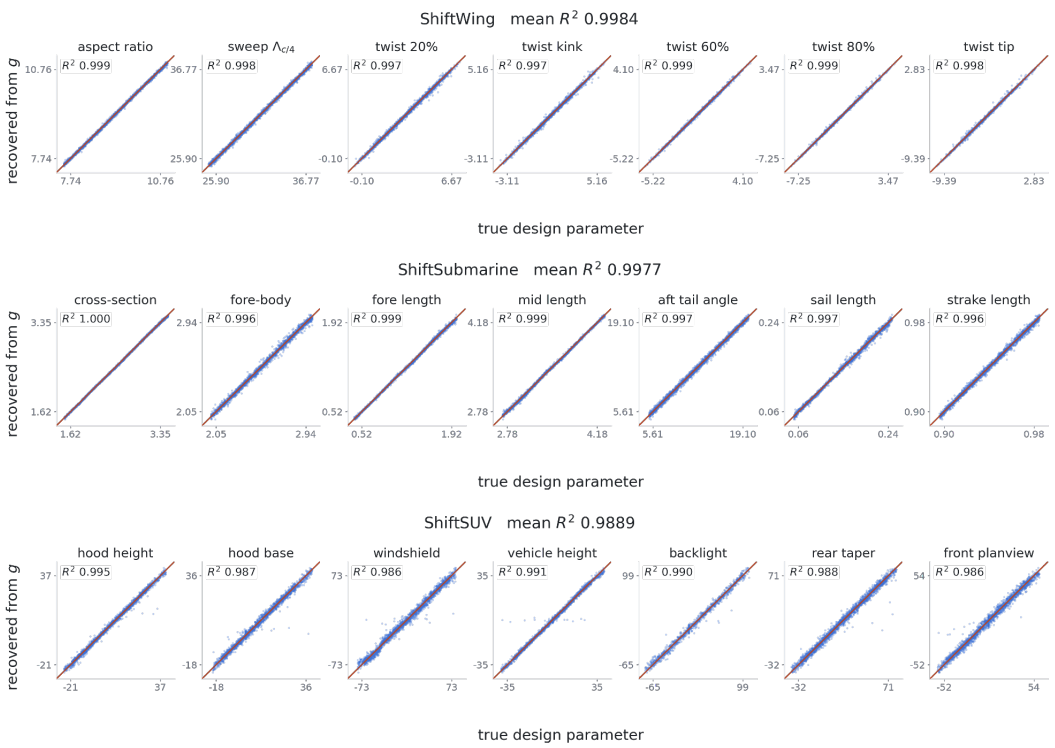


Figure A.4: Design-variable probe recoverability across benchmark datasets. True design parameters are compared with values recovered from the frozen geometry representation using linear probes.

G.2 Continuous operating-condition generalization

Because the geometry representation does not receive the operating condition vector, the same geometry representation can be reused across different flow conditions while the decoder conditions the field prediction on c. On SHIFT-Wing, the evaluated Mach range is 0.50–0.90. Ten Mach states are used during training, while Mach 0.7750 and Mach 0.8125 are held out entirely. Pressure prediction remains above $R^2 = 0.990$ across the evaluated conditions, with $R^2 = 0.9922$ and $R^2 = 0.9944$ at the two held-out Mach numbers. Figure A.5 shows the accuracy across the full sweep.

G.3 Cross-physics decoder transfer

To evaluate reuse across physical response spaces, a pre-trained aerodynamic geometry encoder is frozen and transferred to AutoHood 3D structural displacement prediction, following the frozen-backbone strategy common in parameter-efficient adaptation (Houlsby et al., 2019; Hu et al., 2022). The geometry encoder and its projection layers remain fixed, while only the new physics decoder is trained. The transfer configuration contains 0.56 M trainable decoder parameters, compared with 2.30 M for a model trained entirely from scratch, a $4.1\times$ reduction. Per-epoch results are reported in Table 5 of the main text.

H Differentiable Design Optimization

D-JEPA supports differentiable design optimization because gradients can be propagated from a design objective through the physics decoder and learned geometry representation back to the underlying design variables. Specifically, for design variables θ, the differentiable path is

$$\theta \to G(\theta) \to (P, N) \to E_\vartheta(P, N) = g \to D_{\phi_s}(g, q, c) \to \hat{F} \to J(\hat{F}),$$

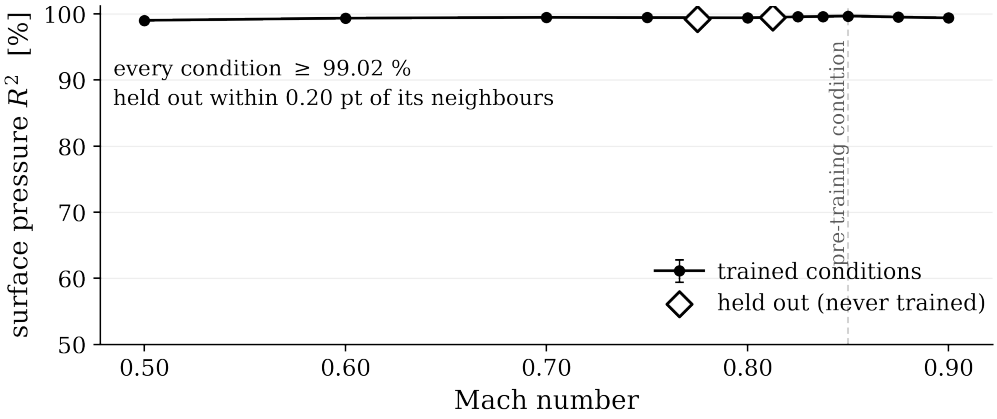


Figure A.5: Operating-condition generalization across Mach numbers on SHIFT-Wing. Surface pressure R^2 is evaluated across trained and held-out Mach conditions using the same geometry representation.

where $G(\theta)$ generates the corresponding geometry and J denotes the design objective computed from the predicted physical field. We evaluate this capability as a downstream application of the learned surrogate rather than as a separate representation-learning objective. Gradients are computed using automatic differentiation through the geometry representation and physics decoder. For SHIFT-Wing, SHIFT-SUV and SHIFT-Submarine, optimization trajectories show strong alignment between the optimized objective and latent search directions, with objective correlation values of approximately $R^2 = 0.99$, 0.99 and 0.98, respectively. To keep optimization within the learned design space, design-variable bounds are imposed through linear design probes together with a k-nearest-neighbor trust-region constraint. The candidate designs produced by this procedure are the ones validated against high-fidelity CFD in Table 6.

I Computational Hardware and Runtime Requirements

The primary hardware configurations and approximate wall-clock runtimes are summarized in Table A.6. Runtime values are provided to contextualize the computational requirements of the reported experiments. For decoder-transfer experiments, freezing the geometry encoder reduces the number of trainable parameters and eliminates backward-gradient storage through the frozen representation. A forward pass through the frozen encoder is still required. Consequently, transfer efficiency arises from both reduced optimizer memory requirements and the removal of backward computation through the geometry encoder.

Table A.6: Primary computational hardware configurations and approximate wall-clock runtimes across pipeline stages.

Execution pipeline stage	**Compute hardware**	**Wall-clock time**
D-JEPA/AeroJEPA training (400 epochs)	1 × NVIDIA T4 (16 GB)	4–8 hours / dataset
Decoder transfer fine-tuning (400 epochs)	1 × NVIDIA T4 (16 GB)	∼2 hours
GeoTransolver baseline (500 epochs)	8 × NVIDIA H100 (80 GB)	∼12 hours / dataset
Mach-condition fine-tuning (200 epochs)	1 × NVIDIA T4 (16 GB)	∼3 hours

The reported runtimes depend on dataset size, solver mesh resolution, batching and hardware utilization and should therefore be interpreted as approximate execution measurements rather than hardware-independent complexity estimates.

J LIMITATIONS

The current evaluation is limited to a small set of parameterized engineering datasets, primarily covering aerodynamic and structural response fields, with cross-physics transfer demonstrated in a single CFD-to-structural setting. Although the results show strong design-variable recoverability, the degree of recoverability varies across geometry families, suggesting that some geometric degrees of freedom may be less explicitly represented in the learned latent space. The operating-condition transfer experiments are also evaluated over a limited range of conditions and do not establish robustness to substantially different flow regimes or boundary conditions. Similarly, the optimization experiments use differentiable surrogate models and predefined objectives within the evaluated design spaces, whereas practical engineering problems may involve constraints, discrete variables, noisy objectives and stronger out-of-distribution conditions. Broader multi-physics benchmarks, more diverse geometry families and constrained real-world optimization therefore remain important directions for future evaluation.